



ЛИНГВИСТИЧЕСКИЕ ТЕХНОЛОГИИ ИДЕНТИФИКАЦИИ ЗНАНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Шаронова Н.В.

*Национальный технический университет
"Харьковский политехнический институт",
г. Харьков, ул. Пушкинская, 79/2, тел. 707–63–60,
e-mail: nvsharonova@mail.ru*

На кафедре интеллектуальных компьютерных систем НТУ «ХПИ» проводятся исследования, целью которых является разработка методов и моделей для интеллектуальной обработки знаний, содержащихся в информационных системах широкого назначения, создание условий для внедрения, поддержания эффективного функционирования и развития самых разных информационных систем. При этом возникает вопрос: а что же такое знания?

В процессе работы над созданием современных компьютерных систем, решающих интеллектуальные задачи (в частности, понимание текстов на естественном языке), на первый план выдвигается проблема представления и извлечения знаний [1-4]. Основой функционирования таких систем является база знаний (БЗ), в которой хранятся сведения о заданной предметной области. Термин «знания» трактуется разными науками по-разному. Поскольку понятие «знания» достаточно сложно, то каждая наука, оперирующая этим понятием, вводит свои частичные его определения, и чаще всего эти определения носят антропоцентрический характер. Эти определения адресованы, прежде всего, человеку, а для построения формальных моделей они неприменимы. Существующее стандартизованное определение понятия «знания» выглядит следующим образом: «Знания – совокупность фактов, отношений, закономерностей и эвристических правил, отображающая уровень осведомленности о проблемах некоторой предметной области» [3].

Приведем еще одно полезное определение. *Знание – это результат адекватного отражения действительности человеком в виде представлений, понятий, теорий, суждений* [2]. Если говорить о знаниях, которые хранятся и обрабатываются компьютерными системами, то в этом случае определение понятия «знания» может быть сформулировано следующим образом: «Знаниями принято называть хранимую (в ЭВМ) информацию, формализованную в соответствии с определенными структурными правилами, которую ЭВМ может использовать при решении проблем по таким алгоритмам как логические выводы» [5].

Из различных определений понятия «знания», ориентированных на использование его при работе с компьютером, можно привести одно из самых удачных: *Знания – это информация, представленная в ЭВМ, которая имеет ряд особенностей: внутренняя интерпретируемость; структурированность; связность; семантическая метрика; активность.*



Если говорить о строго формализованном определении понятия «знания», то в работе [3] приводится следующее определение: *«Знание о факте – это отношение, выраженное некоторым высказыванием. Факт – это действительное состояние всех интересующих нас мест некоторого предметного пространства»*. В этой же работе указывается, что знание не такое определенное понятие, как факт. Оно лишь ограничивает множество возможных состояний мест предметного пространства.

На уровне представления знаний в компьютерной системе отражены как отдельные элементы знаний, так и связи между ними. Уровень представления знаний отличается следующими особенностями: интерпретируемостью, наличием классифицирующих связей, наличием ситуативных отношений [3]. Кроме того, для уровня представления знаний характерны такие признаки, как наличие специальных процедур: обобщение, наполнение имеющихся в системе знаний и т.д. Рассмотрение эволюции проблемы машинного понимания в искусственном интеллекте обнаруживает, что именно разные типы знаний становились краеугольным камнем, методологическим фундаментом компьютерных понимающих систем нескольких поколений.

В недавно опубликованной монографии [1] проведен краткий анализ состояния проблемы в области моделирования идентификации знаний, проанализирована проблема формализации текстов, представленных на естественном языке. Рассмотрены методы интеллектуальной обработки информации: Data Mining и Text Mining. Показана необходимость дальнейшей разработки и исследования математического аппарата алгебры конечных предикатов, метода компараторной идентификации и возможности их применения в моделях представления знаний, системах искусственного интеллекта для формализации различных информационных процессов. Проанализировано и исследовано математическое и лингвистическое обеспечение автоматизированных информационных библиотечных систем. Приведен анализ систем обработки знаний с использованием онтологий.

Список литературы

1. Лингвотехнологии идентификации знаний в информационных системах : монография / О. В. Канищева, Н. В. Шаронова. – Saarbrücken, Deutschland : LAP LAMBERT Academic Publishing, 2013. – 173 с. – На рус. яз.
2. Булкин В.И., Шаронова Н.В. Математические модели знаний и их реализация с помощью алгебропредикатных структур: Монография, НТУ «ХПИ», МЭГИ, Харьков, Донецк, 2010, – 304 с.
3. Бондаренко М. Ф. Мозгоподобные структуры: Справочное пособие. / М. Ф. Бондаренко, Ю. П. Шабанов-Кушнарченко. Том первый. Под редакцией акад. НАН Украины И.В. Сергиенко. – К. : Наукова думка, 2011. – 460 с.
4. Хайрова Н.Ф., Шаронова Н.В. Лингвистические технологии экстракции и идентификации знаний // Тези доповідей Міжнародної науково-технічної конференції "Інтелектуальні технології лінгвістичного аналізу" (м. Київ, 22-23 жовтня 2013 р.). – К. : НАУ, 2013. – С. 7.
5. Осуга С. Обработка знаний / С. Осуга – М. : Мир, 1989. – 293 с.